

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 01-172900

(43)Date of publication of application : 07.07.1989

(51)Int.Cl. G10L 5/02

(21)Application number : 63-290388 (71)Applicant : INTERNATL BUSINESS MACH
CORP <IBM>

(22)Date of filing : 18.11.1988 (72)Inventor : PENN STEVEN C
WILKES MICHAEL D

(30)Priority

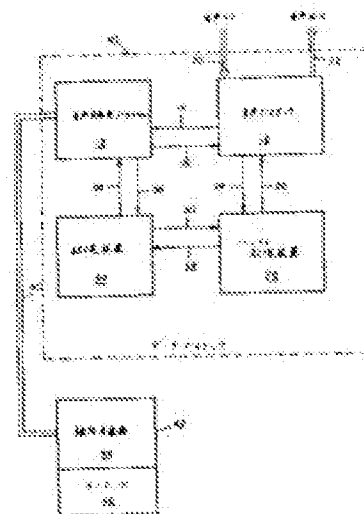
Priority number : 87 135430 Priority date : 21.12.1987 Priority country : US

(54) SPEECH DATA PROCESSOR

(57)Abstract:

PURPOSE: To enable data processing similar to a text editing program by providing a means which converts a received speech recording signal into digital data blocks, a means which records them as a continuous sequence, and a means which displays it.

CONSTITUTION: This speech data processor includes a speech processing system including a circuit 18 which converts a received speech sound record into digital blocks which each represents a corresponding sound record part at specific time intervals and a storage device 34 which stores the said data blocks as the continuous sequence. Further, the data processing system includes a display device 44 for the data blocks which displays a figure based upon the time to a user. Consequently, the speech processing system can be obtained which has an editing function for insertion, deletion, or copying, i.e., similar to a text editing program.



⑬ 日本国特許庁(J P)

⑭ 特許出願公開

⑫ 公開特許公報(A)

平1-172900

⑮ Int. Cl.⁴

識別記号

庁内整理番号

⑯ 公開 平成1年(1989)7月7日

G 10 L 5/02

J-8622-5D

審査請求 有 請求項の数 2 (全12頁)

⑰ 発明の名称 音声データ処理装置

⑱ 特 願 昭63-290388

⑲ 出 願 昭63(1988)11月18日

優先権主張 ⑳ 1987年12月21日㉑ 米国(U S)㉒ 135430

⑳ 発 明 者	ステイブン・クライ グ・ベン	アメリカ合衆国テキサス州ジョージタウン、トンコワ・ト ライル6005番地
㉑ 発 明 者	マイケル・デヴィド・ ウィルケス	アメリカ合衆国テキサス州オースティン、ウエスト・アツプ ルゲイト501番地
㉒ 出 願 人	インターナショナル・ ビジネス・マシーン ズ・コーポレーション	アメリカ合衆国10504、ニวยอร์ก州アーモンク(番地 なし)
㉓ 代 理 人	弁理士 山本 仁朗	外1名

明 細 書

1. 発明の名称 音声データ処理装置

2. 特許請求の範囲

(1) (a) 受信した音声記録信号を、それぞれが予定の期間の該音声記録の対応する部分をあらわすような複数のデジタル・データ・ブロックに変換するための手段と、

(b) 上記データ・ブロックを、上記音声記録信号をあらわす連続的なシーケンスとして記録するための手段と、

(c) 上記データ・ブロックを、時間に基づくグラフィック表示として表示するための手段、

とを具備する音声データ処理装置。

(2) (a) 受信した音声記録信号を、それぞれが予定の期間の該音声記録の対応する部分をあらわすような複数のデジタル・データ・ブロックに変換する段階と、

(b) 時間をあらわす軸を表示する段階と、

(c) 上記時間をあらわす軸に近接して、上記期

間内の上記音声記録信号の相対的な時間位置に対応して上記デジタル・データのグラフィック表示を表示する段階と、

(d) 上記期間内の上記音声記録信号のある時点をあらわす少なくとも1つの、ユーザーが見出すことのできるマークを上記時間をあらわす軸上に表示する段階とを有する、

音声データ表示方法。

3. 発明の詳細な説明

A. 産業上の利用分野

本発明はデータ処理、特にデジタル化されたサウンドの記録に関するデータ処理のアプリケーションに関する。

B. 従来技術

従来は、サウンドトラックの記録はテープレコーダのようなアナログ装置によって行われている。これらのサウンドの記録即ちサウンドトラックの編集には、テープを機械的に接合(splice)するか又はサウンドトラックを第1のテープレコーダから第2のテープレコーダに記録する際に編集し直

す方法が用いられている。

しかしながら、デジタルデータはデジタル処理システムで編集される。デジタルデータ処理に共通する1つのアプリケーションはテキスト処理即ちテキスト編集である。データ処理システムのテキスト編集プログラムは今日では一般に広く使用されている。これらのテキスト編集プログラムはテキスト情報をデジタルデータのシーケンスに変換し、デジタルデータのシーケンスはデータブロックに記憶される。挿入・削除・移動又は複写のような編集機能はこれらのデータブロックを処理することにより容易に実行される。一般に、テキストそのものはテキスト編集プログラムの操作員即ちユーザーに対する表示画面に表示されるので、操作員はテキスト編集プログラムに与えられたコマンドの結果、及びこれらのコマンドに回答して行われるテキスト編集プログラムの活動を観察することができる。

音声サウンドトラックの編集にもテキスト編集プログラムに類似したデータ処理アプリケーション

ンを提供することが望ましい。

C. 発明が解決しようとする問題点

本発明の目的は挿入・削除・移動又は複写のような前記編集機能を具備する音声処理システムを提供することである。

更に本発明の目的は音声データの現在の状態及び音声データに関する音声編集コマンドの動作をユーザーが容易に観察することができる表示装置を有する音声処理システムをユーザーに提供することである。

D. 問題点を解決するための手段

本発明に従って、受取った音声サウンド記録をデジタルデータブロック——各ブロックは所定の時間間隔で対応するサウンド記録部分を表わす——に変換する回路を含む音声処理システムが提供される。サウンド記録を表わすデータブロックを連続シーケンスとして記憶する記憶装置も含まれる。データ処理システムは時間を基準とする図形をユーザーに表示するためデータブロックの表示装置も含む。

本発明に従って、受取ったサウンド記録を表示する方法も提供される。この方法は下記ステップを含む：

(1) 受取ったサウンド記録をデジタルデータブロック——各データブロックは対応するサウンド記録部分を所定の時間間隔で表わす——に変換する。

(2) 時間軸を表示する。

(3) 該時間間隔の対応するサウンド記録部分の相対的な時間位置に一致するようにデジタルデータの図形表示を時間軸に隣接して表示する。

(4) 該時間間隔のサウンド記録の時間位置を表わす時間軸の少なくとも1つの位置を表示する。

E. 実施例

本発明の良好な実施例では、音声編集プログラムはパーソナルコンピュータで実現される。特に本実施例では、IBM PC/AT が使用される。このプロセッサは第1図にデータプロセッサ10として表示される。データプロセッサ10は音声編集プログラム12を含み、音声編集プログラム12は

データプロセッサ10によって実行される。ディスク記憶装置とRAM (ランダムアクセス記憶装置) のどちらも含むことがある記憶装置34、及び記憶装置34と音声プロセッサ18の間のインタフェースとして作用するバッファ記憶装置28も含まれる。バッファ記憶装置28は音声編集プログラム12によって実行されるタスクに適当な大きさの1つの又はひと続きのキャッシュから成ることがある。

音声プロセッサ18はライン20で音声入力を受取りライン22に音声出力を送り出す。良好な実施例では、音声プロセッサ18はワシントン州ブレイン(Brain)のパシフィックマイクロ回路社(Pacific Microcircuits Limited)から市販されているアリエル(Ariel)DSP-16(実時間データ収集プロセッサ)である。この良好な実施例では、音声プロセッサはTI(Texas Instruments)社のTMS 32020 信号プロセッサを含む。音声プロセッサの目的は、ライン20の音声信号入力をデジタル化して音声情報入力を表わすデジタルデー

タを提供すると共に、データプロセッサ10の記憶装置34からのデジタルデータを音声信号に変換することである。変換された音声信号は音声ライン22に出力される。

端末装置42はライン40を介してデータプロセッサ10に接続される。端末装置42は操作卓画面44及びキーボード46から成る。

この音声編集プログラムはIBM ディスプレイライタシリーズの製品の1つのようなテキスト編集プログラムに類似の動作をすることになっている。換言すれば、ユーザーに都合のよいインタフェースが操作員に提供され、操作員は編集タスクを容易に成し遂げることができることになっている。

第1図で、音声編集プログラム12はライン14及び16により音声プロセッサと通信する。音声編集プログラム12はライン38及び36により記憶装置34とも通信する。音声プロセッサ18と記憶装置34の間で渡されるデータはライン24、26、30及び32によりバッファ記憶装置28を通過して渡される。音声編集機能の制御は

ることは当業者には明らかである。しかしながら、どの時点でも一度に表示されるのは2つの前記オブジェクトだけである。

ステップ102に戻って、もしユーザーがプログラムの停止を決めれば、プログラムはステップ108に進み、記憶装置を割振り解除するとともに変更された音声ファイルはどれも保管し、音声編集プログラムを終了する。

任意のサウンドトラック、サウンド記録又は音声情報——音声プロセッサ18（第1図）によって受取られ、デジタル化されてデータブロックになっている——を説明するため、本明細書では用語“音声ファイル”が使用される。良好な実施例では、入力される音声データをサンプリングするのにパルス符号変調技術が使用され、音声情報、例えば50分の1秒間隔（“セグメント”と呼ぶ）の音声情報を表わすデジタル化されたデータブロックを供給する。良好なサンプリング速度は44000サンプル/秒である。サンプルの大きさ及び間隔の長さが変更可能であることは当業者に

音声編集プログラム12で行われる。

第2図は音声編集プログラムの監視ルーチンを示す。ステップ100で、音声編集プログラムが初期設定される。これは任意の開始変数の初期設定及びプログラム実行のための記憶装置の割当てを含む。ステップ102で、プログラムはユーザーが動作を希望するプログラムのモードを決定する。オブジェクト選択モードでは、プログラムはステップ104に進み、オブジェクト指示メッセージを表示しユーザー入力を受取る。オブジェクト指示メッセージは編集されることになっているサウンドトラック、サウンド記録又は音声オブジェクトファイル名の要求である。

オブジェクトファイルを受取ると、プログラムはステップ106に進み、それが有効なオブジェクトであるかどうかを決める。もし有効ではないなら、プログラムはステップ104に戻る。もし有効なオブジェクトが選択されているなら、プログラムはステップ102に戻る。操作員により幾つかのオブジェクトが一度に選択されることがあ

は明らかである。

ステップ102で、もしユーザーがオブジェクトの編集を希望するなら、プログラムはステップ110に進み、オブジェクトが使用可能かどうかを決定する。もしオブジェクトが現在使用可能ではないなら、プログラムは結合子Iに進む。結合子Iは第3図に示すようにステップ112に接続され、音声データベース検索サービスを提供する。この検索機能は編集のため記憶装置からオブジェクトを取出し主記憶装置に入れる。その後、プログラムはステップ114に戻る。同様に、もしステップ110でオブジェクトが使用可能なら、プログラムはステップ114に進む。

ステップ114で、音声データ及び位置リストは表示のため形式化される。音声データ及び位置リストはユーザー又は操作員のための音声編集プログラム情報を含む。

ステップ116で、形式化された音声データ及び位置リストが表示される。初期設定された音声画面及び位置リストの実例を第6図に示す。セク

ション216は音声画面である。セクション228は位置リスト画面である。これらの2つの領域はあとで詳細に説明する。

次に、プログラムはステップ118に進み、ユーザーから有効な要求が入力されているかどうかを決定する。この要求はキーボードコマンドから入力されるか又はマウス入力装置の使用により入力されることがある。もし要求が有効ではないなら、プログラムはステップ116に戻る。しかしながら、もし要求が有効なら、プログラムはステップ120に進む。ステップ120で、プログラムは要求が音声オブジェクトを保管することになっているかどうかを決定する。もしそうなら、プログラムはブロック112(第3図)の音声データベース検索サービスに進み、音声オブジェクトを保管する。もし音声オブジェクトが保管されることになっていないなら、又は音声オブジェクトの保管が完了すると、プログラムはステップ122に進み、該要求はカーソルを移動することになっているかどうかを決定する。カーソルはマウス

によるか又はキーボードのカーソルキーにより移動させることができる。もしカーソル移動が要求されているなら、プログラムはステップ124に進み、画面上のカーソルを移動する。

ステップ126で、プログラムは音声データ又は位置リストが編集されることになっているかどうかを決定する。もしそうなら、プログラムは結合子Pを介して第5図の編集サブルーチンに進む。ステップ128で、プログラムは音声動作が要求されているかどうかを決定する。もしそうなら、プログラムは結合子Oを介して第4図の音声動作サブルーチンに進む。ステップ130で、プログラムはユーザーが現在のオブジェクトの編集からの脱出を要求しているかどうかを決定する。もしそうなら、プログラムは結合子Mを介してステップ102に戻る。さもなければ、プログラムはステップ116に戻る。

第4図は音声動作サブルーチンを示す。ステップ132で、要求された音声動作の初期設定が行われる。ステップ134で、信号プロセッサのプ

ログラムが音声プロセッサ18にロードされる。ステップ136で、該動作を開始するのに必要とする音声データが使用可能であるかどうか決定される。もしそうなら、プログラムは第3図の音声データベース検索サービスに進む。さもなければ、プログラムはステップ138に進み、動作を開始させるコマンドを音声プロセッサ18に送る。

ステップ140で、プログラムはユーザーが中止を要求されているかどうかを決定する。もしそうなら、プログラムはステップ142に進み、音声動作を終了する。さもなければ、プログラムはステップ144に進み、この現在の時点で、あるタイプの動作が要求されているかどうかを決定する。もしそうなら、プログラムはステップ146に進み、動作を実行する適切なコマンドを音声プロセッサ18に送る。次にステップ148で、プログラムはデータを音声プロセッサに又は音声プロセッサから転送することになっているかどうかを決定する。もしそうなら、プログラムは音声データベース検索サービス(第3図)に進む。次の

で、プログラムはステップ140に戻る。

第5図で、音声データ又は位置リストの編集が実行される。ステップ150で、必要な編集のタイプが決定される。もし音声データが編集されることになっているなら、プログラムはステップ154に進む。もし位置リストが編集されることになっているなら、プログラムはステップ152に進む。ステップ152で、プログラムはフィールド入力が正しいかどうかを検査する。ステップ156で、もし該入力が正しいなら、プログラムはステップ160に進み、位置リストを更新する。

ステップ154に戻って、プログラムはデータ編集要求が有効であるかどうかを決定する。ステップ158で、もし要求が有効なら、プログラムはステップ162に進み、表示されている物理的な音声データを再形式化する。ステップ162から、プログラムは第3図の音声データベース検索サービスを経てステップ164に進み、必要に応じて位置リストを更新する。もしステップ156及び158でそれぞれ入力が正しくなく要求が有

効ではないなら、又はステップ160及び164でそれぞれ位置リストの更新が完了すると、プログラムはステップ168に進み、どのフィールドも変化分だけ更新する。

第6図は初期の音声編集プログラム画面を示す。オブジェクト名、例えば“SAMPLER.AUD”が領域202に記載される。現在の音声編集プログラム状況、例えば“編集/停止”が領域204に記載される。音声の長さ（を表わす時間）が領域208に記載され、時間は領域210のセグメントの項にも記載される。オブジェクトのセグメント位置は領域212に記載される。対応するセグメントは領域214にも記載される。現在位置は常に編集すべきブロックの前にあることに注目されたい。音声編集プログラムの現在の版は領域206に記載される。以上は見出し領域200を構成する。音声データは領域216に現われる。2つの時刻線224及び226の表示は分及び秒を刻むマークを含む。第1の時刻線224は第1のオブジェクトの時刻線を表わす。第2の時刻線226は第

2のオブジェクトに使用される。ポインタ218は現在のオブジェクトの時刻線を示す。

ポインタ220及び222も領域216に含まれる。これらのポインタはオブジェクトの開始及び終了を表わす。現在の例にはオブジェクトがないので、開始及び終了を示すポインタはどちらも時刻0を指す。

領域228に表示される位置リストは“ID”セクション230、対応する（固定又はスライド）位置セクション232、時刻セクション234、タイプセクション236及びパラメータ/注記セクション238から成る。コマンド指示メッセージ領域240も提供され、ユーザーが使用しうる可能性があるコマンドを表示する。

第7図で、“SAMPLER.AUD”と呼ばれる新しい音声オブジェクトファイルが記録により生成中である。波形250は全体の音声記録の一部分である。このオブジェクトファイルの長さは音声の長さセクション258の記載のように5分36.58秒である。時間軸の終了マーク252はこの終了点を示

す。位置リストで、開始マーク254は開始時刻が0であることを示す。終了マーク256はこの記録の長さが5分36.58秒であることを示す。

終了点を表示するためボリューム歴波形250は画面移動されている。“SAMPLER.AUD”の記録の5分36.58秒は16829個のデータセグメント（1セグメントは1/50秒）に等しい。ボリューム歴は表示間隔のセグメントの平均ボリュームである。各ボリューム歴に示されたセグメントの表示間隔は音声編集プログラム画面を提供するのに使用された操作卓の分解能により異なる。

波形250のボリューム歴は全てのアプリケーションで適切であるとは限らない。それゆえ、波形250の基準として別の特性、例えば周波数が用いられることがある。

位置リストでは、開始点は固定位置を持つものとして記載される。固定位置はこのマークの位置が時間軸に対し永続的に固定されることを表わす。それに対し、終了マーク256はスライドするマークである。これはこのマークの位置は時間軸全

域で変化するがサウンド記録の或る部分に対しては相対的に一定であることを表わす。したがって、“SAMPLER.AUD”オブジェクトファイルに追加セグメントが付加されると、終了点はそれに応じて外方にスライドされる。開始マーク254及び終了マーク256のタイプはどちらもPマークとして記載される。Pマークは永続的であるのでユーザーはそれを除去することはできない。

第8図では、もうひとつの音声オブジェクトが選択されている。音声オブジェクト標識260は上部の音声オブジェクト線から下部の音声オブジェクト線に移動されている。下部の音声オブジェクトは“FINALE.AUD”と記載され、35.00秒（即ち1750個のデータセグメント——1セグメントは1/50秒）のサウンド記録から成る。これは更新された位置リスト——開始ポインタ267及び終了ポインタ268から成る——でも表示されている。

第9図では、声オブジェクトファイル“FINALE.AUD”の全体が、オブジェクトファイル“SAMPLER.AUD”の全体が、オブジェクトファイル“SAMPLER.AUD”の全体が、オブジェクトファイル“SAMPLER.AUD”の全体が、

AUD"の末尾に挿入されている。これはポインタ277の位置変更により指示された最初の音声オブジェクトの再選択を生ずる。波形274 (FINALE, AUD 波形)は位置270のあとに波形276として挿入されていることに注目されたい。時間位置1 (参照番号270)は古い "SAMPLER.AUD"が終了した場所を示すために付加されている。位置1は現在 5分 38.58秒 (SAMPLER.AUD の古い末尾)と記録され、Sマークとして表示される。Sマークは後に操作員により使用されることがある事象の発生を指示するため音声編集プログラムにより自動的に生成されるシステムマークである。終了マークも時間列に示すように更新されている。

かなりの編集の結果を示す第10図ではユーザーは SAMPLER.AUDオブジェクト内で適切な位置を付加しそれらの特性を位置リストで定義している。SAMPLER.AUD オブジェクトの波形282は画面移動により開始位置に戻されている。ユーザーに示された画面は 2分40秒の波形を表示する。しかしながら、オブジェクトはもっと長いことがよくあ

るので、ユーザーは関心のある部分を見るため表示画面を移動することができる。波形282は時間位置2 (281)、時間位置3 (280)、時間位置4 (284)及び時間位置5 (285)のような幾つかの時間位置を含む。これらの時間位置の指定は位置リスト286に含まれる。位置2 (281)は、位置リスト286に従ってボリュームが減少されることになっている SAMPLER.AUDと混合される予定の、もうひとつの音声オブジェクト "VOICE 1.AUD"の開始を指定するために付加されている。

その結果、(位置リスト286の)時間位置2は SAMPLER.AUDトラックのボリュームコマンド、混合コマンド、及び混合すべきトラックのボリュームコマンドを含む。位置3は音声再生の間に実時間で行われる生の会話(talkover)の始めにユーザーマーク(Uマーク)として付加されている。位置4は特殊な音響効果を指定する。位置4で混合される音響効果はパラメータ/注記セクションで再指定されることに注目されたい。位置3及び

4はどちらもスライドするので、編集動作にかかわらず関連する音声部分とともに留まることができる。換言すれば、これらの位置は適切な音声データとともに一定しているが、音声オブジェクトが編集されると時間軸に沿って"スライド"することができる。位置5は位置2に関連する動作と類似のもうひとつの"画面外の声"(voice-over)の動作である。位置7はあとで付加することができるもうひとつの"画面外の声"の箇所を指定する。終了位置は表示されないが、必要なら画面移動により観察することができる。

第11図では、記録機能が呼出され現在の表示画面のマーク8とマーク3のあいだに50秒の無音を記録したあとの表示画面がユーザーに示される。これは場所286における波形の欠如により明白である。音声編集プログラムは記録された無音をマーク3の前に挿入し、この新しい記録の開始位置を明確にするためマーク8を付加している。音声分割がなされると、スライド特性を有するマーク(例えば、マーク3及び4)は新たにデータ

が記録されるのに先立って移動させられていることに注目されたい。固定特性を有するマーク(例えば、マーク5)は同じ時間変位を示す位置に留まっている(即ち、それらの位置は時間軸に関して一定である)。位置リスト287が終了ポインタを表示していないことにも注目されたい。オブジェクトファイルの波形のように、位置リストは上方又は下方に画面移動して全体のリストの異なる部分を表示することができる。

F. 発明の効果

前述のように、本発明は音声データの現在の状態及び音声データに関する音声編集コマンドの動作をユーザーが容易に観察することができる表示装置を有する音声処理システムをユーザーに提供する。

4. 図面の簡単な説明

第1図は音声編集プログラムを含むデータ処理システムのブロック図、

第2図は音声編集プログラム監視ルーチンの流れ図、

第3図は音声編集プログラム検索サブルーチンの流れ図。

第4図は信号プロセッサに対する音声編集プログラムインタフェースの流れ図。

第5図は音声編集プログラムの編集サブルーチンの流れ図。

第6図は音声編集プログラムの最初の表示画面を示す図。

第7図は1本のサウンドトラックを表わす音声編集プログラム表示画面を示す図。

第8図は2本のサウンドトラックを表わす音声編集プログラム表示画面を示す図。

第9図は複写機能を表わす音声編集プログラム表示画面を示す図。

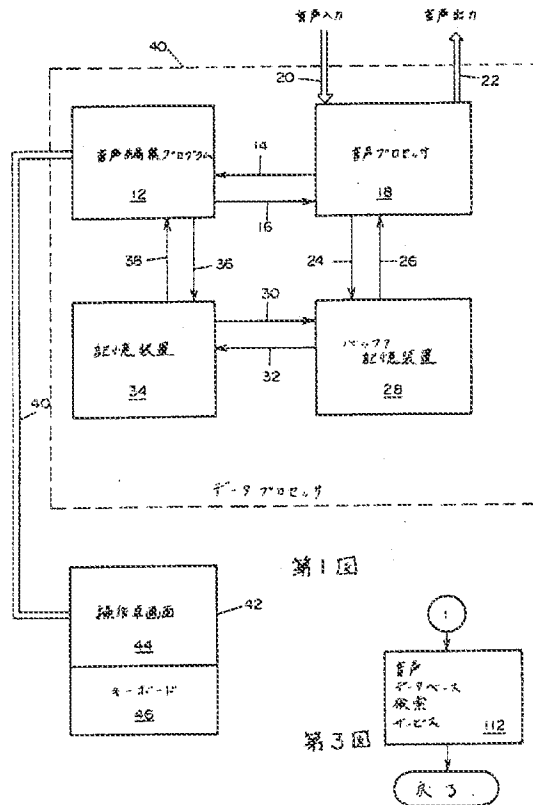
第10図は時間マークの使用を表わす音声編集プログラム表示画面を示す図。

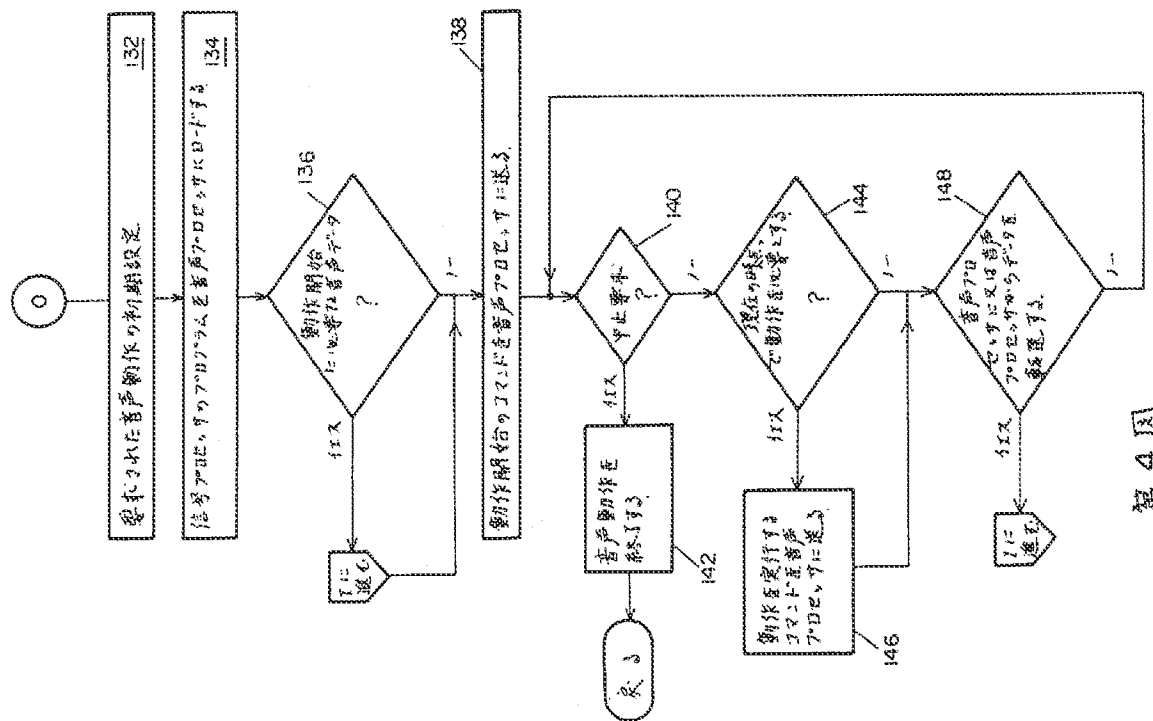
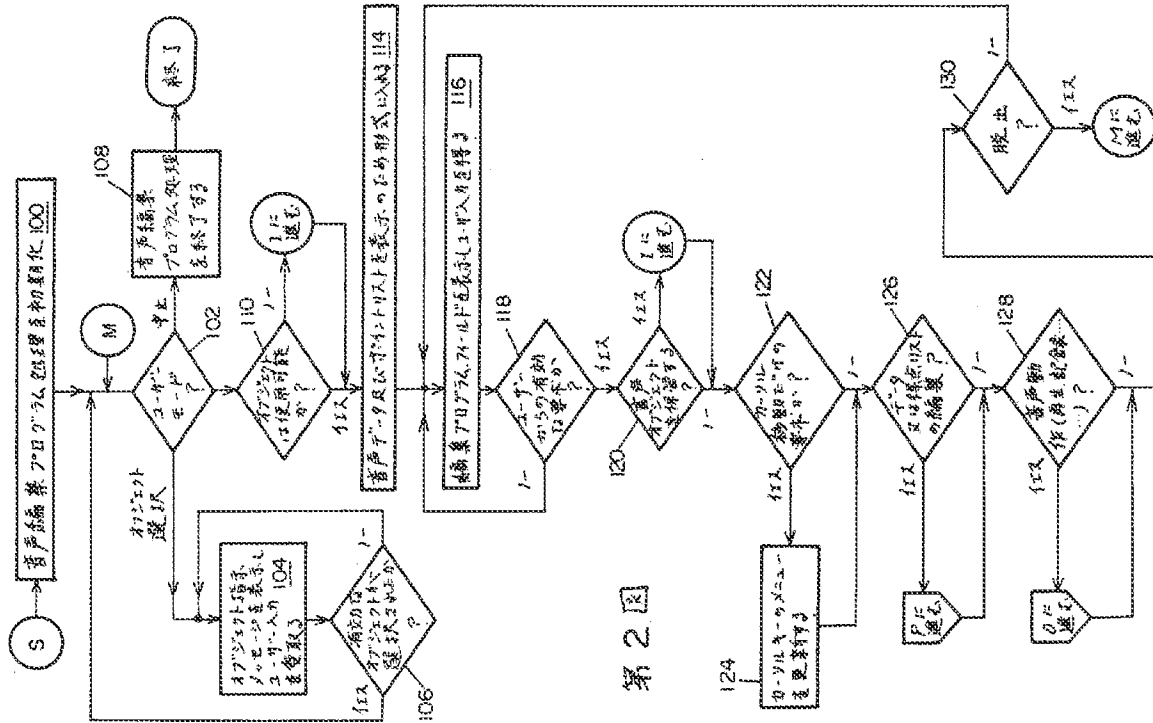
第11図は記録機能を表わす音声編集プログラム表示画面を示す図である。

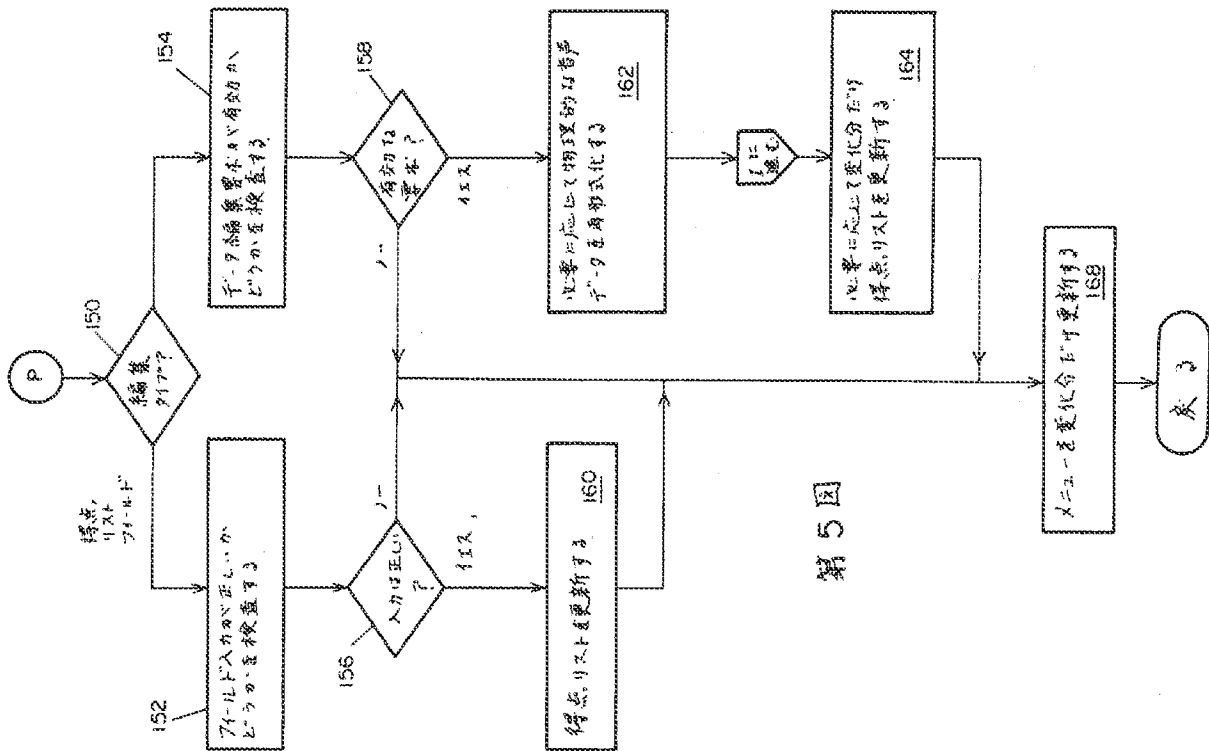
10・・・データプロセッサ、12・・・音声編集プログラム、18・・・音声プロセッサ、28・・・

バッファ記憶装置、34・・・記憶装置、42・・・端末装置、44・・・操作卓画面、46・・・キーボード。

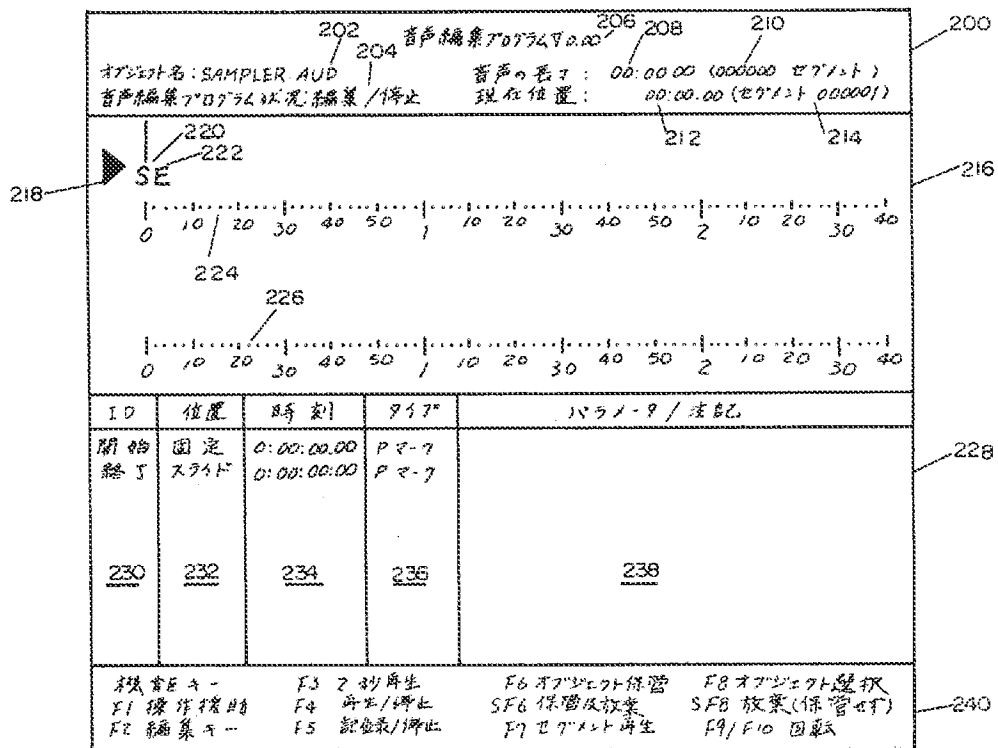
出願人 インターナショナル・ビジネス・マシーンズ・コーポレーション
代理人 弁理士 山本 仁 朗
(外1名)



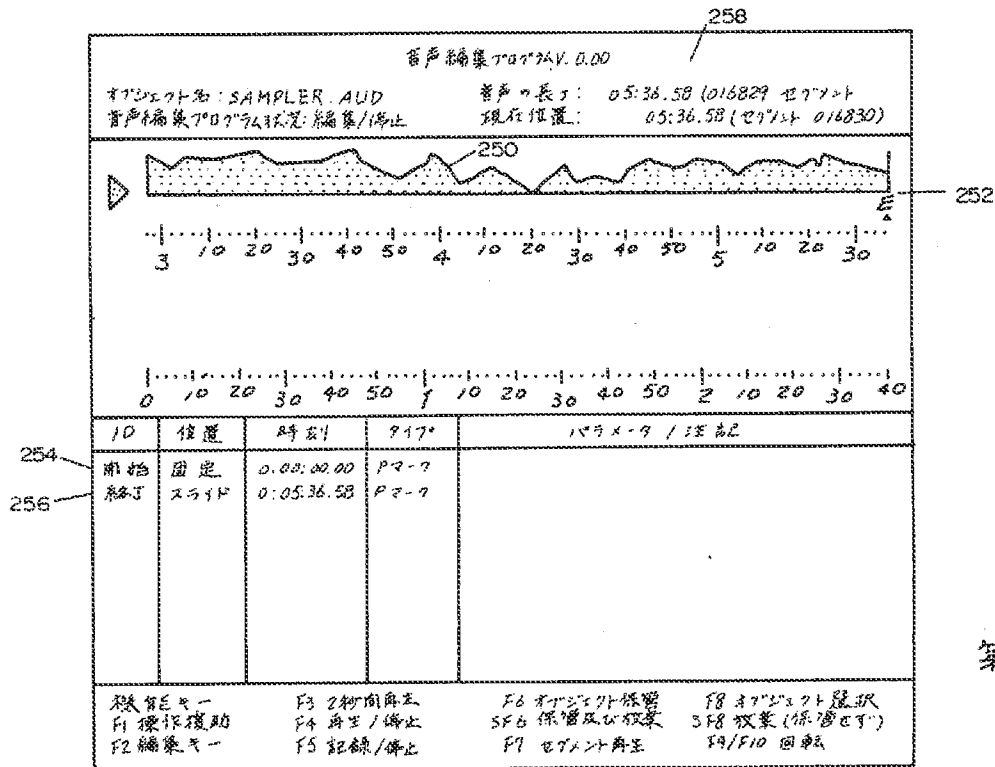




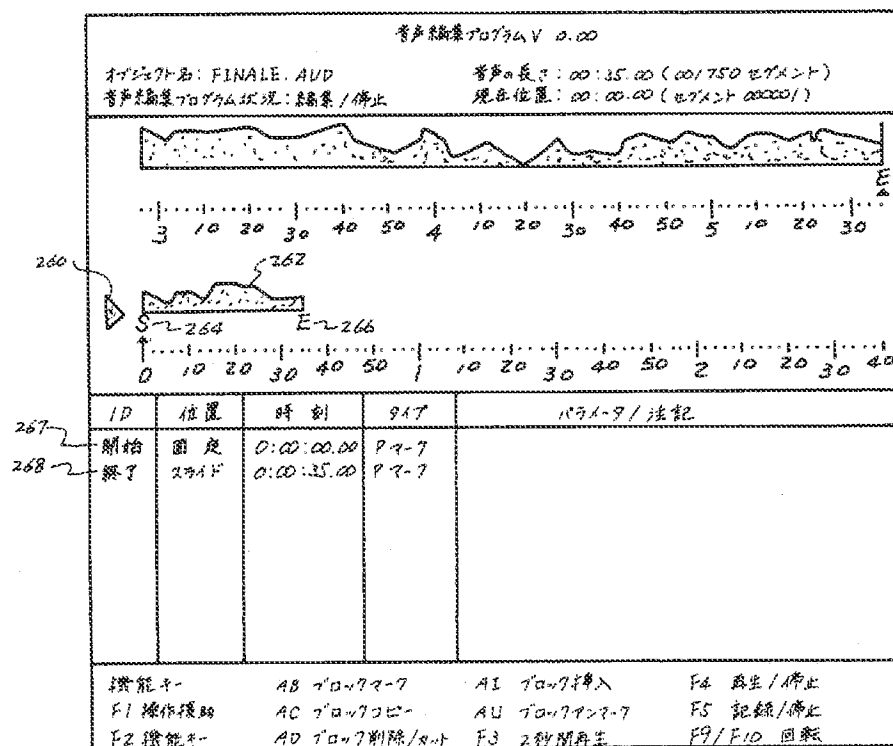
第5図



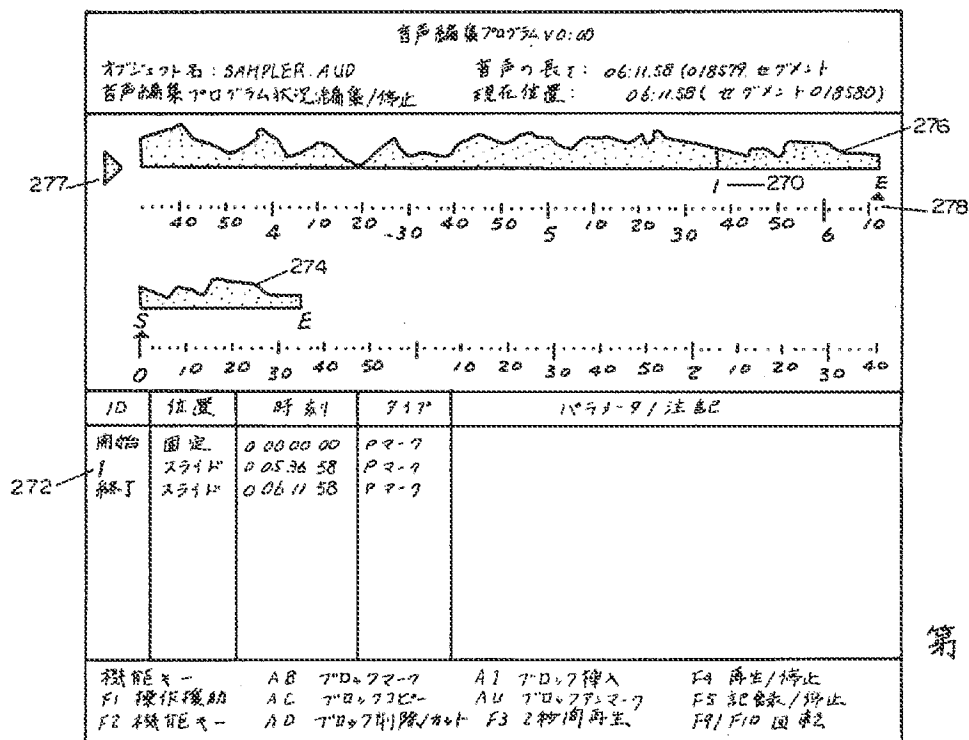
第6図



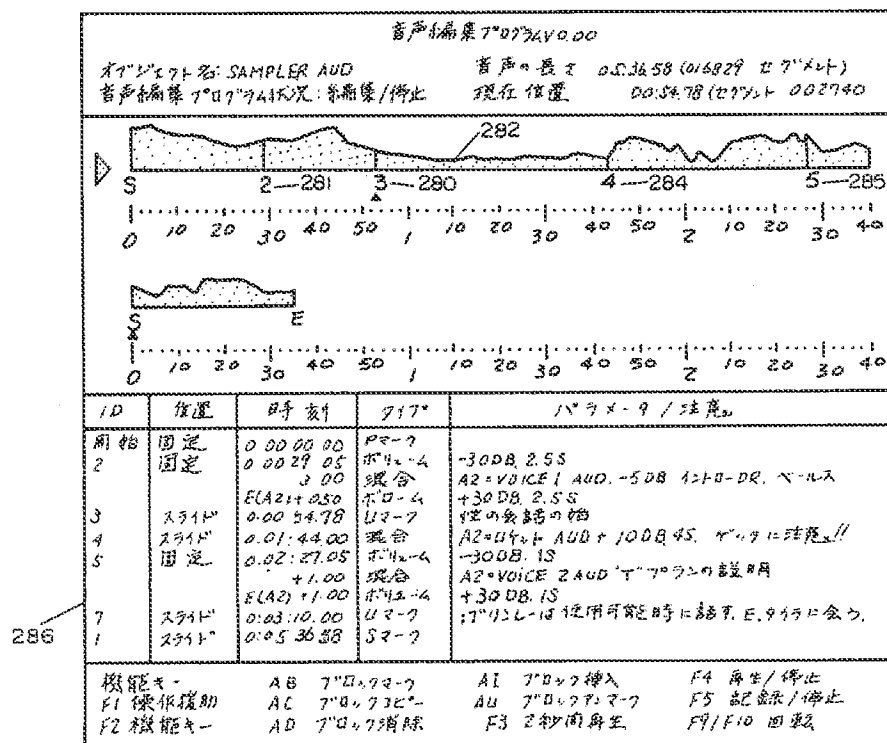
第7図



第8図



第9図



第10図

